

## The Research Data Centre of the German Federal Employment Agency: Data Supply and Demand between 2004 and 2009

Heining, Jörg

Veröffentlichungsversion / Published Version  
Arbeitspapier / working paper

### Empfohlene Zitierung / Suggested Citation:

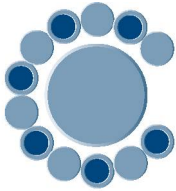
Heining, J. (2009). *The Research Data Centre of the German Federal Employment Agency: Data Supply and Demand between 2004 and 2009*. (RatSWD Working Paper Series, 129). Berlin: Rat für Sozial- und Wirtschaftsdaten (RatSWD). <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-75356-2>

### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:  
<https://creativecommons.org/licenses/by/4.0/deed.de>

### Terms of use:

This document is made available under a CC BY Licence (Attribution). For more Information see:  
<https://creativecommons.org/licenses/by/4.0>



German Council for Social  
and Economic Data (RatSWD)

[www.ratswd.de](http://www.ratswd.de)

# RatSWD

## *Working Paper Series*

Working Paper

No. 129

The Research Data Centre of the  
German Federal Employment Agency:  
Data Supply and Demand between  
2004 and 2009

---

Jörg Heining

---

October 2009

---

## Working Paper Series of the Council for Social and Economic Data (RatSWD)

---

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

*RatSWD Working Papers* are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

# The Research Data Centre of the German Federal Employment Agency: Data Supply and Demand between 2004 and 2009\*

Jörg Heining

October 20, 2009

## Abstract

The Research Data Centre (FDZ) of the German Federal Employment Agency (BA) in the Institute for Employment Research (IAB) was founded in 2004 and is intended mainly to facilitate access to BA and IAB micro data for non-commercial empirical research using standardised and transparent access rules. Five years after its foundation the FDZ is acknowledged as producer of innovative data products and has become an visible element of the research community. Furthermore, linkages to other data producers and the internationalization of data products and data access have been promoted by the FDZ.

## 1 Introduction

The Federal German Employment Agency (Bundesagentur für Arbeit, BA) is the most important public producer of labour market data in Germany. Although the amount of micro data with high research potential was huge, the transfer of these data to external researchers was rather limited before 2004. Data privacy regulations and especially the lack of standardised ways of data access constrained the dissemination of these data to the research community. In order to facilitate data access several attempts have been made by the Institute for Employment Research (IAB) in the 1990s. But none of these attempts succeeded in transparent and systematic ways of data access for external researchers.

In 2001 the Commission to improve the informational infrastructure by cooperation of the scientific community and official statistics (Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik, KVI) recommended the foundation of so called research data centres to public producers of microdata in Germany (Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik 2001). The

---

\*For many helpful comments and discussions I would like to thank Stefan Bender, Benedikt Hartmann and Peter Jacobebbinghaus. I also would like to thank Jörg Paulsen (IAB, Documentation and Service) for evaluating the literature database. All errors and shortcomings are - of course - mine. Jörg heining Institute for Employment Research (IAB), Tel.: +49-(0)911-179-5392, Fax: +49-(0)911-179-1728, Joerg.Heining@iab.de



BA followed this recommendation in spring 2004 and founded the Research Data Centre (FDZ) of the BA in the IAB. First co-funded by the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) the FDZ is since 2006 completely financed by the BA after a positive evaluation of the German Council for Social- and Economic Data (Rat für Sozial- und Wirtschaftsdaten). Besides the FDZ ten other research data centres have been established by now.<sup>1</sup>

Since its foundation the FDZ has a clear perception of itself (Kohlmann 2005). It is primarily a service-oriented institution, providing access to high-quality micro data under compliance of the current data privacy legislation. Furthermore, the members of the FDZ also conduct own (empirical) research with and on the data. Only by dealing with the content and the options to analyse the available datasets a sound knowledge about the quality of the data and their research potential can be accumulated. Without this knowledge neither high-level advisory service nor continuous improvements of the data quality are possible. All activities in the dimensions of data access, the development of new data products and research are not restricted to the BA or to national borders. Openness, networking and an international focus are also part of the FDZ's self-perception.

Five years after its foundation the FDZ is regarded as one of the most important supplier of high-quality micro data on the German labour market. Up to the moment, 10 datasets covering various aspects of the labour market have been developed, documented and are offered to the scientific community. Most of the available datasets stem from the notification process of the social security system, internal procedures of the BA and/or surveys conducted by the BA or the IAB. Moreover, the FDZ has started several projects to enlarge its data supply by trying to link the existing data to other administrative data sources or surveys, both nationally and internationally. One of these projects already succeeded in the Panel WeLL. Here a employee survey for the project "Further Training as a Part of Lifelong Learning" has been linked to the data of the BA. Another project aimed to enlarge the existing data supply and to facilitate accessibility is the "Research Data Centre in Research Data Centre" (RDC-in-RDC) approach, also promoted by the FDZ. The basic idea of this approach is that the data of other producers of (administrative) micro data may be accessed through the FDZ and vice versa, even abroad.

It is no surprise that due to their quality the data products of the FDZ are highly in demand by national and international researchers. In 2008, for example, 141 new contracts enabling data access for external researchers have been concluded. Another aspect explaining the high demand for the data products is that the data products of the FDZ are comparatively easy to access. Depending on the degree of anonymisation, three ways of data access, i.e. scientific use files, remote data access and on-site use, are available. But the quality and the accessibility of the data may explain the high demand only in parts. Even more interesting for researchers is the high research potential of the data which already led to numerous high-ranked publications.

After five years of successful work, the aim of this paper is to take stock and to provide an

---

<sup>1</sup>An overview of German research data centres is provided on the website of the German Council for Social- and Economic Data <http://www.ratswd.de/eng/>.

overview over recent activities and future developments of the FDZ. The paper is organised as follows: Section 2 describes the main tasks of a research data center or rather the FDZ. Data supply and the ways of data access are covered in Section 3 while the demand for the data products of the FDZ is outlined in Section 4. Future developments and research activities are reported in Section 5 and finally, Section 6 concludes.

## 2 Main Tasks

As described in the introduction, the FDZ perceives itself not solely as an institution providing data access, but as a mediator between the producers and the users of data. Beside the provision of transparent and standardised ways of data access under compliance of the current data privacy regulations, the following tasks are inherited by the FDZ (Bender et al. 2009c):

- **Development and implementation of data models**  
 Providing access to the microdata of the BA does not mean to transfer raw data to external researchers. The FDZ develops and updates specific data products with high research potential and carries out the necessary preparations and adjustments of the raw data. In order to prevent disclosure and the violation of data privacy regulations, the development and application of anonymisation strategies are tasks of the FDZ.
- **Documentation of the data and individual counseling**  
 In order to assess and tap the full research potential of the data researchers have to know about both technical aspects of the data generation and the statistical properties. Therefore, detailed documentation of the data and individual counseling have to be provided by the FDZ.
- **Promotion of the data**  
 The FDZ promotes the spread of information on the data and data access by the organisation and active participation in scientific conferences, workshops or seminars.
- **Active research on and with the data**  
 Doing active research on and with the data is essential to generate knowledge about the data. Only by being an active member in the research community, the members of the FDZ are able to follow the current methodical discussions in order to assess the research potential of the data and to provide high-level individual counseling. Therefore, the FDZ has not only to fulfill the role as a service provider, but also as an active research unit.

## 3 Data Supply and Data Access

### 3.1 Data Supply

The datasets offered by the FDZ are generated from three different sources: the notification process of the social security system, internal procedures of the Federal Employment Agency and surveys (see Figure 1).

Data from the notification process of the social security process are submitted in accordance with the German Data and Transmission Act (Verordnung über die Erfassung und Übermittlung von Daten für die Träger der Sozialversicherung, DEÜV) and contain information on employment and unemployment as covered by the social security system. Information on establishments, participation in measures of active labour market policy, payment of social benefits and job search stem from process data internally produced by the BA. Survey data are either used to amend data from the other two sources or stand as datasets for themselves.

At the moment, the FDZ offers external researchers 10 different data products which may be distinguished by establishment data, individual/household data and integrated establishment and individual data. A short description of this datasets is given in the paragraphs below whereas detailed information are provided on the website of the FDZ (<http://fdz.iab.de/en.aspx>).

#### 3.1.1 Establishment Data

**IAB Establishment Panel (IABB)** The IAB Establishment Panel (Kölling 2000, Bellmann 2002 or Fischer et al. 2009) is an annual representative survey on establishments in Germany on various topics such as the determinants of labour demand. It provides information on up to approx. 16,000 establishments per year and has been conducted by the IAB since 1993 in West Germany and since 1996 in East Germany. Each establishment includes at least one employee subject to social security contributions at the time of the interview (June 30th each year). Some standard topics like the development of employment, business policy and development, investments, et cetera are queried annually. Additionally, each wave focuses on some special topics, for example future demand for qualified employees in 2007 or job security and locational security of industries in 2006.

**IAB Establishment History Panel (BHP)** The Establishment History Panel (BHP) (Spengler 2008) is composed of yearly cross sectional datasets up to 1975 for West-Germany and 1992 for East Germany. Every cross section contains all establishments in Germany with at least one employee liable to social security on June 30th. Since 1999 also establishments with no employees liable to social security but with at least one marginal part-time employee are included. The cross sections can be combined to form a panel. The BHP contains information about the branch of industry and the location of the establishment. Furthermore there is the number of employees liable to social security as well as marginal part-time employees, both in total and subdivided

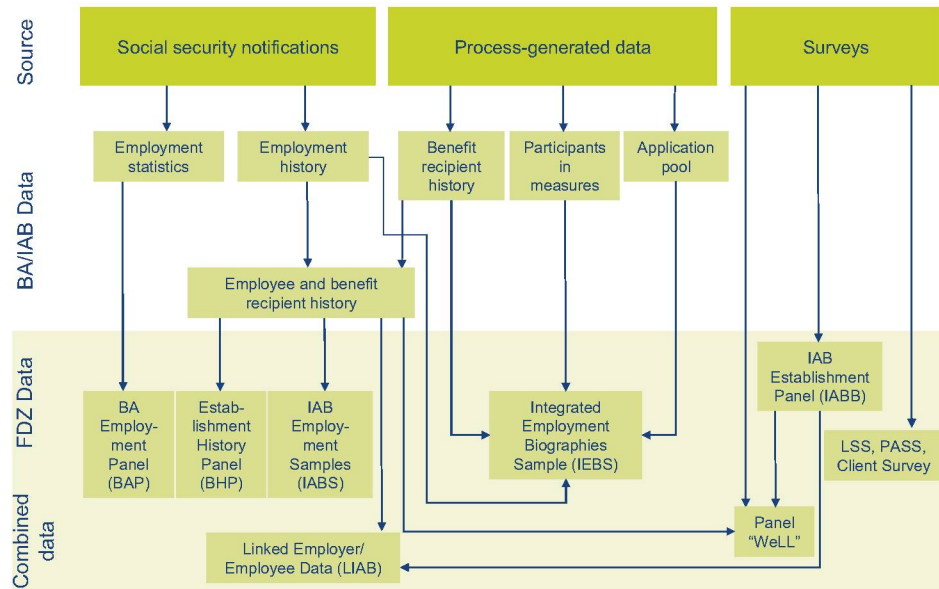


Figure 1: Data Sources and Paths.



Table 1: Versions of the IAB Employment Samples (IABS).

	Scientific use file		Weakly anonymised version
	Basic file	Regional file	
Period	1975 - 1995	1975 - 1997 1975 - 2004	1975 - 2004
Strong Differentiation on	Professions, industries, cooperate workforce	Regions	Professions, industries, regions

into by gender, age, occupational status, qualification and nationality (since 1999). Quartiles of ages and wages are also given, both for full-time employees only as well as for all employees.

### 3.1.2 Individual/Household Data

**BA Employment Panel (BAP)** The BA Employment Panel (BAP) (Koch and Meinken 2004) is generated from the quarterly employment statistics of the BA since 1998 and is annually updated by the IAB. The employment statistics cover all employees in Germany who are subject to social security contributions. The database is intended mainly to deliver up-to-date information, therefore, belated employer notifications will not be considered. A 2% sample is drawn with the reference date being the last day of each quarter. Apart from information on employees, the BAP has details on unemployment and participation in measures of active labour market policy. Additionally, marginal part-time employment has been included since 1999. Finally, the data contain a few characteristics on establishments. The BAP may especially be utilized as a basis for cross-section as well as time-series, panel and cohort analyses. Moreover, the BAP represents the official statistics of the BA, while data from the IABS (see next paragraph) are different due to the different editing procedures. Similar to the IABS, the BAP contains few variables on establishments in addition to the characteristics of individuals.

**IAB Employment Samples (IABS)** The IAB Employment Samples (Bender and Haas 2002) are 1% or 2% samples drawn from the longitudinal processed database of employment notifications to the social security system and supplemented by information on benefit recipients. Since the IAB Employment Samples are updated subsequently to the BAP database, even belated employer notifications are considered. The samples cover a continuous flow of data on employment as well as on receipt of unemployment benefits, unemployment assistance and maintenance allowance; therefore, they are highly suitable for performing analyses on the employee and benefit recipient history. Depending on the way of data access and therefore on the degree of anonymisation, the covered time period as well as the differentiation of contained variables may differ. Table 1 gives an overview over the various versions of the IABS.

**Integrated Employment Biographies Sample of the IAB (IEBS)** Similar to the IABS, the Integrated Employment Biographies Sample (IEBS) (Jacobebbinghaus and Seth 2007) is based on the employee history and the benefit recipient history compiled in the IAB. By way of contrast, the IEBS integrates other available data sources too, such as the participants in measures and applicant pool databases. Thus, the IEBS embraces event history data on employees liable to social security, benefit recipients, persons who are searching for employment, unemployed persons and participants in measures of active labour market policy. The IEBS allows for more detailed overviews of employment histories as well as comprehensive analyses on active labour market policy. The spells related to participations in measures contain numerous cases on the following measures in active labour market policy: job-creation measures, settling-in allowance, business start-up allowance, measures to promote vocational training, German language courses and more. In its current version the IEBS covers the period from 1990 to 2008.

**Cross-sectional survey "Life Situation and security 2005" (LSS 2005)** The Cross-Sectional Survey "Life Situation and Social Security 2005" (LSS 2005) was carried out by the Institute for Applied Social Sciences (infas) commissioned by the IAB between November 2005 and March 2006 (Meßmann et al. 2008). 20,832 persons affected by the German Hartz IV reform were interviewed. These persons comprise recipients of unemployment or social assistance of employable age (15-64) who were registered as unemployed or job seekers in December 2004 and employable persons in need of assistance aged 15 to 64 who were registered as unemployed or as job seekers in January 2005 and who were recipients of Unemployment Benefit II (ALG II) according to the German Social Code Book II (SGB II). The specific characteristic of the LSS is that it represents a unique dataset which allows for drawing up analyses of the referenced target groups directly after the amalgamation of unemployment and social assistance alongside the introduction of the German Social Code Book II (Hartz IV reform).

**Panel Study "Labour Market and Social Security" (PASS)** The Panel Study "Labour Market and Social Security" (PASS) (Trappmann et al. 2009a and b) is a novel dataset in the field of labour market, welfare state and poverty research in Germany. PASS is a new central source for drawing up analyses with reference to the labour market and poverty situation in Germany as well as to the situation of recipients of benefits in accordance with the German Social Code Book II. The applied survey design is a two-stage random sample including 300 postal code areas drawn in a first step, from which household communities in joint receipt of benefits taken from the register data and households from a residential buildings sample are drawn in a second step. The survey units consist of two partial populations: persons and households in receipt of Unemployment Benefit II and persons and households registered as residents of Germany. Initially, a personal interview was carried out with the heads of all selected households. Subsequently, members aged 15 or older were interviewed. Persons aged 65 or older were presented with an abridged questionnaire referred to as a pensioner's questionnaire. PASS covers a broad spectrum of research questions with regard to the issues of employment and

unemployment and is suitable to provide even very detailed information. The panel also includes large-scale socio-demographic characteristics and subjective indicators such as contentment, fears and problems or employment orientation. With almost 19,000 interviewed persons in more than 12,500 households PASS is currently one of the most comprising panel surveys in Germany. The survey is conducted annually. The first wave was carried out by TNS Infratest Sozialforschung between December 2006 and July 2007, commissioned by the IAB. Since 2009 the second wave is available.

**Client Survey on Organizational Forms of SGB II-Agencies 2007/08** The Client Survey on Organizational Forms of SGB II-Agencies 2007/08 was carried out by the Centre for European Economic Research (ZEW), Mannheim, the Institute for Work, Skills and Training (IAQ), University of Duisburg-Essen, and TNS Emnid, Bielefeld. The Swiss Institute for Empirical Economic Research (SWE) acted as an advisor. When the SGB II was introduced in 2005, the tasks to be carried out in association with assisting the individuals concerned were also changed. A standard model was not developed for this assistance, but instead different forms of organisation (agencies responsible for implementing the SGB II) compete with each other and are to be evaluated according to the law (§ 6c SGB II). To this end a large association of research institutions conducted extensive surveys among the agencies responsible for implementing the SGB II and among the clients of these agencies. The unusual feature of the client survey which was developed in this project is that the details about the respondents are linked with a wealth of information about the agency (organisation, strategies, labour market indicators) (Oertel et al. 2009). In addition, detailed questions are asked about activation measures implemented by the agencies and about the results of these measures for the recipients of benefits in accordance with SGB II. Furthermore, the survey also includes detailed questions about the living situations and employment histories of the benefit recipients, as well as an instrument to measure employability. The survey is representative of the 154 selected agencies responsible for the implementation of SGB II and covers 24,422 individuals.

### 3.1.3 Integrated Establishment and Individual Data

**Linked Employer-Employee Data from the IAB (LIAB)** The Linked Employer-Employee Data from the IAB (LIAB) (Alda et al. 2005) allow for simultaneous analysis of the supply and demand side of the German labour market. For this purpose the IAB Establishment Panel data are matched with personal data generated in labour administration and social security data processing. Two different models of the LIAB data are available. The LIAB cross-sectional model which contains both information on individuals and data from the IAB matched on a specific reference date (June 30th). The LIAB longitudinal model is available in three versions. In every version of the LIAB longitudinal model establishments is matched to the employment and benefit receipt history of the employed individuals. Depending on the version, the covered periods and establishments may differ.



**Panel WeLL - Employee Survey for the project "Further Training as a Part of Lifelong Learning"** For the project "Further Training as a Part of Lifelong Learning" (WeLL) a small establishment survey was conducted and in addition the employees of the surveyed establishments were interviewed by telephone about their further training activities. This survey is designed as a longitudinal survey with three survey waves conducted in 2007, 2008 and 2009. The first wave of the panel WeLL was conducted by the Institute for Applied Social Sciences (infas) between autumn 2007 and January 2008. For this about 6400 employees from 149 establishments were interviewed. The survey design is a two-step sample for which first a sample of establishments was drawn which was stratified according to the criteria of establishment size, sector, location, further training activities and investment activities. In the second step the respondents were drawn from the total of employees in these establishments. Owing to this specific survey design the sample is not representative of German establishments and employees but is specifically designed for analysing in-firm further training activities. As part of the project "Further Training as a Part of Lifelong Learning" (WeLL) an extensive innovative dataset is being established which contains information about further vocational training from employers and employees. In a first step the FDZ is now going to provide a scientific use file for external users (Bender et al. 2009a). This dataset contains the data from the first wave of the employee survey supplemented by some establishment information. In this survey the further training activities and the employment biographies of the respondents from the beginning of 2006 up until the time when the survey was conducted were recorded in detail. In addition, socio-demographic characteristics and information about income, household, job satisfaction and expectations regarding the future were collected.

### 3.1.4 Documentation

Closely connected to data supply is its documentation. Without a clear and detailed documentation of the data empirical research will not be successful. The FDZ publishes two publication series (in German, English translations are available for most of the issues) in order to provide users and interested researchers with descriptions of the data and methodological aspects when dealing with the data. The FDZ Datenreport series contains data documentations and information on statistical aspects of the datasets. Methodological problems and aspects are addressed in the FDZ Methodenreport series. The two publication series as well as working tools for the specific datasets may be accessed free of charge through the FDZ website <http://fdz.iab.de/en.aspx>.

## 3.2 Data Access

All the individual-level and establishment data offered by the FDZ are subject to data protection legislation. The legal basis for data access is mainly § 67 of the German Social Code Book X and § 282 section 7 of the German Social Code Book III. In compliance with this legal regulations, three ways of data access are offered as described below. When granting external researchers



access to the micro data of the BA there is a conflict of interest between the maintenance of the research potential in the data on the one hand and in the disclosure of individual information worth of protection on the other hand. Therefore, the FDZ follows the principle that the need for protection determines the ways and the limits of data access. In general, the more detailed the data the more restrictive is the data access offered by the FDZ.

Table 2 provides an overview over the various datasets of the FDZ and the ways to access them.

### 3.2.1 Scientific Use Files

Scientific use files are factually anonymous datasets, which are provided to scientific institutions within the scope of § 282 section 7 of the German Social Code Book III. Scientific use files may solely be used for research only. Data access for teaching or commercial interests is not possible. In order to get access to scientific use files offered by the FDZ, researchers have to send an application to the FDZ. Beside the names of the project members, the time frame of the project as well as the relation of the project to labour market research has to be specified. Moreover, the modes of ensuring confidentiality by organisational and technical measures have to be described. After the application is approved by the FDZ, a contract is signed and the researcher will receive the data in the required format (SAS, Stata, SPSS or ASCII). Any attempt of deanonimisation by the researcher or merging the data with other data is strictly prohibited. Scientific use files are available for the Regional File 1975 - 2004 of the IAB Employment Sample (IABS)<sup>2</sup>, the Integrated Employment Biographies Sample of the IAB (IEBS), the survey "Life Situation and Social Security 2005" (LSS 2005), the panel "Labour market and Social Security" (PASS) and the Client Survey. Since September 2009 the scientific use file of the WeLL data is also offered to the scientific community. All scientific use files are available free of charge.

### 3.2.2 Remote Data Access

Beside scientific use files the FDZ offers external researchers to access the data remotely. By definition, remote data access<sup>3</sup> means that the researcher is able to access data from his home desktop computer via a secure internet connection at anytime. Program files for the evaluation of the data may be started by the researcher and information on the status of the job are provided automatically. In order to prevent disclosure an automatic check of the output files is performed before the files are sent to the researcher or can be downloaded. Unfortunately, due to technical and legal constraints this "ideal" way of remote data access has not been implemented by any of the German research data centres so far. Instead, remote data execution is applied. This means that syntax files (in SAS, Stata or SPSS) that have been sent by external researchers are processed with the original data by the FDZ staff. After that the FDZ staff screens the

---

<sup>2</sup>Only the Regional File 1975 - 2004 of the IABS is distributed by the FDZ. The Basic File 1975 - 1995 (ZA-Nr. 3136) as well as the Regional File 1975 - 1997 (ZA-Nr. 3348) is distributed by GESIS - Leibniz Institute for the Social Sciences <http://www.gesis.org>.

<sup>3</sup>A definition of remote data access as well of remote data execution is given in Hundepool et al. 2009.

output files, preserves anonymity and sends them to the researcher. In order to guarantee a smooth and quick procession the FDZ provides detailed information, for example codebooks, questionnaires, and syntax files for dealing with special problems of the data on its website. Additionally, artificial test data are offered for familiarisation with the data and for preparation of the syntax files.

The procedure to get data access via remote data access is comparable to scientific use files. Externals researchers have to send a form that includes information on the applicant, the aim of the analyses and their content to the FDZ which decides on the application. Compared to scientific use files, remote data access has the disadvantage that an external researcher is not able to directly control the procession of his syntax files with the data and to receive his outputs immediately. But remote data access provides access to less anonymised data, which may offer a higher research potential than scientific use files. The FDZ charges no fee for this way of data access. Remote data access is available for IAB Establishment Panel (IABB), the Establishment History Panel (BHP) and the IAB Employment Panel (BAP). For the IAB Employment Sample (IABS, weakly anonymised version), the Integrated Employment Biographies (IEBS, weakly anonymised version) and the Linked-Employer-Employee-Data (LIAB), remote data access is only available after on-site use.

### 3.2.3 On-site Use

The third way of data access offered is on-site use, i.e. direct access to the data via a stay as a guest researcher in the FDZ. This way of data access provides direct access to only weakly anonymised datasets where researchers may dispose of the full information on the content of the variables (only identification variables are deleted) on request. In accordance with § 75 of the German Social Code Book X, researchers who wish to visit as guest researchers have to apply for on-site use in written form. Beside the applied dataset(s), the application has to include the names and the affiliation of the project members, a technical and a non-technical description of the project and the relation of the research project to the German social security system has to be outlined. Furthermore, it must be depicted why the particular research project is of public interest. After the request has been approved by the FDZ, the Federal Ministry of Labour and Social Affairs (Bundesministerium für Arbeit und Soziales, BMAS) has to be asked for authorisation of the data access. Data access is finally granted after signing a contract between the researchers and the FDZ. The maximum duration of a research visit is two weeks. Data protection is guaranteed by a special configuration of the guest researchers' workplace and disclosure reviews at the end of each research visit.

### 3.2.4 Access for Non-German Users

Access to the data products of the FDZ is not restricted to German researchers or researchers living in Germany. Non-German researchers are granted data access under the same terms and conditions as for Germans and/or resident researchers. Though there is no formal discrimination

Table 2: Data products and ways of data access.

Dataset	On-site use	Remote data access	Scientific use file
<b>Establishment Data:</b>			
IAB Establishment Panel (IABB)	✓	✓	
IAB Establishment History Panel (BHP)	✓	✓	
<b>Individuell Data/Household Data:</b>			
IAB Employment Samples (IABS)	✓	✓*	✓
BA Employment Panel (BAP)	✓	✓	✓
Integrated Employment Biographies Sample of the IAB (IEBS)	✓	✓*	✓
Cross-sectional survey "Life Situation and Social Security 2005" (LSS 2005)			✓
Panel Study "Labour Market and Social Security" (PASS)			✓
Client Survey on Organizational Forms of SGBII-Agencies 2007/08			✓
<b>Integrated Establishment and Individual Data:</b>			
Linked Employer-Employee Data from the IAB (LIAB)	✓	✓*	
Panel WeLL - Employee Survey for the Project "Further Training " as a Part of Lifelong Learning"			✓

\* Only available after on-site use.

Table 3: Number of new contract according to dataset and data access. Source: Bender et al. 2009a, p. 9.

<b>Dataset</b>	<b>Data access</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2005-2008</b>
IABB	Remote data access	29	36	35	42	142
	On-site use*	2	8	7	12	29
BHP	On-site use*	0	2	7	7	16
IABS	On-site use*	1	7	8	8	24
	Scientific use file	31	26	19	17	93
BAP	Remote data access	1	3	9	6	19
	On-site use*	0	1	1	0	2
	Scientific use file	8	0	4	4	16
IEBS	On-site use*	0	3	3	2	8
	Scientific use file	0	0	0	4	4
LSS	Scientific use file	0	0	0	3	3
PASS	Scientific use file	0	0	0	9	9
LIAB	On-site use*	9	14	21	27	71
<b>Total</b>		<b>81</b>	<b>100</b>	<b>114</b>	<b>141</b>	<b>436</b>
	Scientific use file	39	26	23	37	125
	Remote data access	30	39	44	48	161
	On-site use	12	35	47	56	150

\* On-site use is followed by remote data access.

of researchers from abroad, in respect of visits of the FDZ they are disadvantaged due to higher travelling expenses. For this reason a grant to assist researchers from abroad was set up by the FDZ in 2007. When a request is approved, the FDZ grants partial funding of the accommodation expenses for guests from non-German-speaking countries. In 2008, seven guests made use of this service. The establishment of this grant was evaluated positively by the German Council of Science and Humanities (Wissenschaftsrat) in its report (Wissenschaftsrat 2007, p.55).

## 4 The Demand for Data

Five years after its establishment the FDZ has an excellent reputation in the scientific community. It is widely acknowledged as an innovative producer of high-quality micro data on the German labour market and its data products enjoy great popularity nationally as well as abroad. Between the period 2005 and 2008, 436 contracts were concluded. In this period the number of new contracts increased continuously, starting with 81 contracts in 2005 and reaching 141 in 2008. The IAB Establishment Panel (IABB) experienced the highest demand, while the IAB Employment Sample (IABS) was runner-up (see Table 3).

Most of the contractual partners and guest researcher at the FDZ are researchers living in Germany. But there is also a significant part of users coming from abroad. In the period from 2006 to 2008 the numbers of contractual partners as well as on-site users from abroad increased



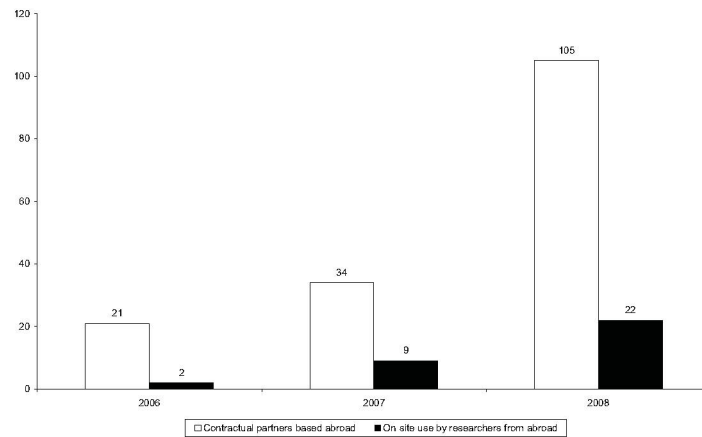


Figure 2: Contractual partners based abroad and on-site use by researchers from abroad from 2006 to 2008. Source: Bender et al. 2009a, p. 19.

as demonstrated by Figure 2.

#### 4.1 Scientific Use Files

After a decline in 2005 and 2006 the demand for scientific use files reached nearly its initial level from 2005 (see Table 3). In 2008, 37 projects have been given data access via scientific use files by the FDZ. Causal for the increase in 2008 was the availability of the new scientific use files for the Integrated Employment Sample (IEBS) and the surveys PASS and LSS. The decline in new projects for the scientific use files of the IAB Employment Samples and the BA Employment Panel may be explained by the fact that projects generally last over one year. Therefore, researcher who for example successfully applied for the IAB Employment Sample in one year, will not apply for the dataset in the next years.

#### 4.2 Remote Data Access

Though remote data access has the disadvantage for the researchers of not directly seeing the data, it is a widely accepted way of data access. The number of cases of remote data access have grown steadily over the years, reaching 1,390 at the end of 2008 (see Table 4). Remarkable is the increase in the number of cases between 2006 and 2007. Due to a growing number of researchers accessing the data of the FDZ by remote data access the case numbers more than doubled. The case numbers stayed nearly constant in 2008 although the number of projects using remote data

Table 4: Number of cases of remote data access by dataset. Source: Bender et al. 2009a, p. 11.

<b>Dataset</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>Total</b>
BAP	1	13	55	65	134
IABB	196	257	368	431	1252
BHP	-	-	78	79	157
IABS	-	26	79	173	278
IEBS	6	11	34	51	102
LIAB	155	370	713	591	1829
<b>Total</b>	<b>359</b>	<b>677</b>	<b>1328</b>	<b>1390</b>	<b>3754</b>

Table 5: Total time required for verification of evaluations using remote data access (in hours). Source: Bender et al. 2009a, p. 11.

<b>Dataset</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>
IABB	55,8	219,1	140,4	157,9
BAP	1,0	11,5	29,5	15,2
BHP	-	-	24,1	71,2
IABS	-	8,9	29,3	26,9
IEBS	-	2,8	10,4	8,9
LIAB	24,0	149,7	149,4	121,7
<b>Total</b>	<b>80,8</b>	<b>391,9</b>	<b>383,2</b>	<b>401,7</b>

access or on-site use followed by remote data access rose. This development may be explained by a learn effect on part of the users who gained more experience in handling the data. Broken down into figures for individual datasets, the LIAB data and the IABB were the two data sets most responded to.

After the syntaxes have been processed to the data the outputs are screened by the FDZ staff. In conjunction with the growing number of cases also the required time for verification has grown. In 2008 the FDZ staff spent over 400 hours screening outputs to avoid disclosure of individuals and to ensure data privacy regulations (see Table 5). One should recognize that beside the number of cases the scope and the complexity of the specific evaluations also influence verification time.

### 4.3 On-Site Use

In order to enable external researchers the access to weakly anonymised datasets via on-site use the FDZ is equipped with one guest room and five workstations. Like for the other ways of data access the numbers of researchers accessing data via on-site use considerably increased since the establishment of the FDZ. Starting with 22 visits by external researchers in 2005, the number of visits reached 256 in 2008. In total 482 visits by external researchers have been registered during

Table 6: Publications produced using the FDZ data by type of literature. Reference date: 16.01.2009. At least one author per publication is a user of the FDZ and no author is affiliated to the IAB in the years of publication. Source: Evaluation by the department "Documentation and Service" of the IAB.

	IABB	BHP	BAP	IABS	IEBS	LIAB	Total
SSCI Journals	11	0	0	12	3	1	27
Other refereed Journals	15	1	3	11	5	7	42
Other Journals	12	0	0	5	1	3	21
Working Paper/Discussion Paper	21	5	7	59	21	31	144
Monographs	65	1	9	14	3	4	96
Articles in collected editions	4	0	2	4	0	1	11
<b>Total</b>	<b>128</b>	<b>7</b>	<b>21</b>	<b>105</b>	<b>33</b>	<b>47</b>	<b>341</b>

the period 2005 to 2008. Though this huge increase in visits the existing facilities at the FDZ were sufficient to allow every researcher enough time for his guest stay. The average duration of visit was three days in 2005 to 2007 before declining to two days in 2008 .

#### 4.4 Publications

The success of the FDZ and its importance for the scientific community are not only documented by the high demand on data products of the FDZ as outlined in the sections above. It is also demonstrated by the publications produced using data from the FDZ (see Table 6). From 2004 to 2008 the users of the FDZ have produced a total of 341 publications. 27 of these contributions were published in SSCI (Social Sciences Citation Index) journals and a further 42 in refereed journals. Even more publications in SSCI or refereed journals can be expected in the near future due to the large the number of working/discussion papers .

## 5 Infrastructural Developments and Research Activities

The FDZ works on and innovates its infrastructure continuously in order to optimize ways of data access and data supply. Moreover, the FDZ conducts own research on and with its data. The aim of this section is to illustrate both the current/future infrastructural developments and the research activities of the FDZ. In the first part, current or future projects concerning the infrastructure are described. Part two focuses on the research activities, including the development of new datasets.

### 5.1 Infrastructural Developments

The first project concentrates on the conception and establishment of a metadata (management) system on international standards. The data are only of use to researchers if they are documented

and if information on the data generating process is available. To strengthen the capacity of the FDZ and to facilitate exchange of statistical information, the FDZ in cooperation with the IT department of the IAB (ITM) is developing a metadata specification on international standards.

Another project focusing on the infrastructural aspects is JoSuA (Job Submission Application)<sup>4</sup>. The FDZ is implementing a more efficient system to run remote data access called JoSuA which was developed by the Institute for the Study of Labor (Institut zur Zukunft der Arbeit, IZA). As mentioned above, the ideal way of remote data access could not be implemented by the FDZ so far. JoSuA is a first step towards this ideal conception of remote data access. With JoSuA the researcher is able to submit syntaxes either by email or using a web interface and to start the evaluation by her-/himself. There are no delays because the job was sent at the weekend or at night and the status of the job can be controlled at anytime. After the output files have been checked by the FDZ staff, they may be downloaded from the web.

The last and the most prestigious of the current projects is the so called "Research Data Centre in Research Data Centre" (RDC-in- RDC) approach. This project aims to facilitate data access and moreover, to extend the available data supply at the FDZ. As described above, less anonymised datasets may be only accessed by (costly) guest stays in Nuremberg which are followed by remote data access afterwards. In order to facilitate guest stays for researchers, the FDZ developed the RDC-in-RDC approach. The basic idea is to offer researchers the possibility to access the data of the FDZ via a secure internet connection from locations different than Nuremberg. In cooperation with the Research Data Centres of the statistical offices of the Länder, on-site use of the FDZ's data shall be offered at initially four different sites. Vice versa, the micro data of the statistical offices of the Länder may be accessed through a guest stay at the FDZ. The RDC-in-RDC approach is technically implemented by special separated work stations at the respective research data centres which are connected to the FDZ in Nuremberg. Researchers apply for data access as described above and additionally specify their preferred location. After approval the FDZ in Nuremberg provides online access to the data. The data themselves never leave the FDZ and the produced output files are saved on a FDZ server. As at a regular guest stay, the output files are controlled by the Nuremberg staff before they are sent to the researcher. It is important to notice that though the data of the FDZ and the statistical offices of the Länder may be accessed through one location, the respective data are stored in different locations. Moreover, it is strictly forbidden to combine the data of the FDZ and the data of the statistical offices of the Länder without special permission.

In order to enable more researchers from North-America to access the data of the FDZ, a branch of the FDZ at the Institute for Social Research (ISR) of the University of Michigan in Ann Arbor is planned. Though there are no differences in the technical implementations and the application process between the "German" and the international RDC-in-RDC approach, the legal situation is different. Most important, only anonymised data are allowed to be accessed from abroad within the RDC-in-RDC approach. Therefore on-site staff is required to specifically

---

<sup>4</sup>A short description of JoSuA is given in Schneider 2009 and Schneider and Wolf 2008.



construct anonymised datasets for approved projects.

## 5.2 Research Activities

The FDZ's research activities are well documented through its publication record. In 2008 FDZ employees published 23 research articles. These include one publication and an editorship in top scientific journals listed in the Social Sciences Citation Index (SSCI) and four publications in refereed journals<sup>5</sup>. The "Conference on the Analysis of Firms and Employees (CAFÉ 2006)", which was held by the FDZ in 2006, was documented both in a special edition of *Labour Economics* and in a refereed NBER book which was published in 2008. This picture is completed by numerous presentations held in Germany and abroad about the research activities conducted at the FDZ. In addition, the FDZ is involved in a number of externally funded projects, co-financed by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG), the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) or the Leibniz Association and carried out in cooperation with universities, research institutes or the Federal Statistical Office. Each of these externally funded projects also included funding for staff.<sup>6</sup> As for the infrastructural developments, only a short overview over some of the current projects is given here.

September 2007 marked the launch of the project "Combined Firm Data for Germany" (KombiFiD)<sup>7</sup>, the second project to be mentioned here. The aim of this project is to link for the first time company data from the statistical offices, the Deutsche Bundesbank and the BA/IAB. The project will run for three years in cooperation with the Research Data Centre of the Federal Statistical Office and the Department for Empirical Economic Research of Leuphana University of Lüneburg. On receipt of a written agreement from the companies, their data will be merged. The next goal is to make data from the FDZ of the BA and the Research Data Centre of the Federal Statistical Office available to scientists for research purposes in early 2010. Another goal of the project is to create a legal framework for permanent linkage of the individual data. The KombiFiD project is pioneering and makes it possible to study economic processes in more detail and more thoroughly than was the case with the data material previously available. The project is funded by the BMBF.

The second project to be mentioned in this context is the construction of a longitudinal-biographical data set, based on the data of the German Pension Insurance, the BA and the IAB called BASiD (Biographical data of selected social insurance agencies in Germany) (Budzak et al. 2009). Both the Research Data Centre of the German Pension Insurance (FDZ-RV) and the FDZ offer longitudinal individual-level datasets. These datasets contain on the one hand information of the social security notifications and on the other hand characteristics of the administrative procedures of both institutions. In each institution only information for the accomplishment

<sup>5</sup>See Bender et al. 2009b for a complete list.

<sup>6</sup>A detailed overview is given in Bender et al. 2009b.

<sup>7</sup>A detailed description of the KombiFiD project is given in Brandt et al. 2008 or Hethey and Spengler 2009 or on the KombiFiD website [http://fdz.iab.de/en/FDZ\\_Projects/kombifid.aspx](http://fdz.iab.de/en/FDZ_Projects/kombifid.aspx).

of their own tasks is kept. For example information concerning the number of children and birth dates is stored only at the German Pension Insurance, whereas information about training measures of the BA is only available at the IAB. The project's intention is to construct a combined data set and to offer the data to researchers as scientific use file and via on-site-use.

Another current project of the FDZ focuses on data on transnational commuting in the German-Danish border region. The IAB Northern Germany Regional Office in cooperation with the Department of Border Region Studies of the University of Southern Denmark is building up a dataset which enables for the first time the analysis of transnational commuting between Germany and Denmark in both directions. By using the method of record-linkage, transnational commuters are identified both in the German and Danish data (see Buch et al. 2009). Together with the FDZ the project partners work on a dataset containing individual employment biographies, which will be accessible via the FDZ. Beside nationality and sex the data will also contain information on payment, industry codes and firm size therefore allowing structural analyses of transnational mobility.

While the projects mentioned so far focused on the development of new datasets, the emphasis of the last research project lies on the evaluation of already existing data. In cooperation with the RWI Essen (Rheinisch-Westfäl- isches Institut für Wirtschaftsforschung e.V) data of the BA's Employment History aggregated on the zip code level are merged with the German Socio-Economic Panel (G-SOEP)<sup>8</sup>. The projects aims at an empirical application of models of social interactions to the patterns and determinants of human capital formation. It is an often observed empirical regularity in Germany that agents who belong to the same group tend to behave similarly and to display similar outcomes with regard to the transition from school to work and career dynamics. In order to evaluate this data on a comparatively small spatial level, for example zip-code level, are required.

## 6 Conclusion

With the foundation of an own research data centre in 2004 the BA followed a recommendation of the KVI in order to facilitate and standardize access to micro data on the German labour market. Just one year after the FDZ has started its work, numerous researchers had applied successfully for data access which impressively demonstrated the necessity and the usefulness of such kind of institution. Therefore, Kohlmann in 2005 already concluded that the setup of the FDZ was a step "into the right direction" (p. 444).

sNow, five years after its foundation, this assessment has definitely turned out to be true. The FDZ has become one of the most important suppliers of micro data on the German labour market. It has been positively evaluated twice (by the Council for Social- and Economic Data (Rat für Sozial- und Wirtschaftsdaten) in 2006 and by the German Council of Science and Humanities (Wissenschaftsrat) in 2007) and acquired an excellent national and international

---

<sup>8</sup>See <http://www.diw.de/soep>.

reputation. Beside the comparatively easy ways of data access and the high research potential of the data, elaborate documentation and excellent individual counseling were responsible for the development in the demand for the data.

The activities of the FDZ in the past five have not been restricted to the administration of the existing data products. Several new data products have been developed and offered to the scientific community. Among these new developments were for example the Establishment-History Panel (BHP) and the scientific use files of the Integrated Employment Biographies (IEBS) or the Panel WeLL. Conferences and Workshops have been held in order to deepen the knowledge on the available data and to expand the network within the scientific community.

The past five years have been a very successful period for the FDZ. But still the development of the FDZ has not come to an end. After the data resources of the BA have been used for the development of (new) data products, the next step for the FDZ will be the linkage of these data with other German micro data as it is intended in the KombiFD or in the BASiD project. Also the further opening of the FDZ to the international data market is one of the aims for the near future. The RDC-in-RDC approach and the establishment of a FDZ branch in Ann Arbor, Michigan will mark another milestone in the development of the FDZ in particular and for the access to administrative German micro data in general.

## References

- [1] Alda, H., Bender, S., Gartner, H.: The linked employer-employee dataset created from the IAB Establishment Panel and the process-produced data of the IAB (LIAB). *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*. 125(3), 327 – 336 (2005)
- [2] Bellmann, L.: Das IAB-Betriebspanel, Konzeption und Anwendungsbereiche. *Allgemeines Statistisches Archiv*. 86(2), 177 – 188 (2002)
- [3] Bender, S., Haas, A.: Die IAB-Beschäftigtenstichprobe. In: Kleinhenz, G. (ed.) *IAB-Kompodium Arbeitsmarkt- und Berufsforschung*, pp. 3 – 12. Nürnberg (2002)
- [4] Bender, S., Fertig, M., Görlitz, K., Huber, M., Schmucker, A.: WeLL - Unique Linked Employer-Employee Data on further Training in Germany. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*. (2009b) *forthcoming*.
- [5] Bender S., Hartmann, B., Haug K., Herrlinger, D., Schmucker, A.: FDZ Annual Report 2008. FDZ Methodenreport. No. 04/2009. (2009a)
- [6] Bender, S., Himmelreicher, R., Zühlke, S., Zwick, M.: Improvement of Access to Data Set from the Official Statistics. RatSWD Working Paper Series. Working Paper No. 118 (2009c)
- [7] Brandt, M., Oberschachtsiek, D., Pohl, R.: Neue Datenangebote in den Forschungsdatenzentren - Betriebs- und Unternehmensdaten im Längsschnitt. *Wirtschafts- und Sozialstatistisches Archiv. AStA*. 2(3), 193 – 207 (2008)



- [8] Buch, T., Niebuhr, A., Schmidt, T.: Cross-border commuting in the Danish-German Border region - Integration, institutions and cross-border interaction. *Journal of Borderlands Studies*. 24(2), 38 – 54 (2009)
- [9] Budzak, U., Hochfellner, D., Steppich, B., Voigt, A.: Das Projekt BASiD: Biographiedaten ausgewählter Sozialversicherungsträger in Deutschland. (2009) *forthcoming*
- [10] Fischer, G., Janik, F., Müller, D., Schmucker, A.: The IAB Establishment Panel - things users should know. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*. 129(1), 133 – 148 (2009)
- [11] Hethey, T., Spengler, A.: Combined Firm Data for Germany (KombiFiD). *Matching Process-Generated Data and Survey Data. Historical Social Review*. 34(3), 204 – 214 (2009)
- [12] Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Longhurst, J., Schulte Nordholt, E., Seri, G., De Wolf, P.: Handbook on Statistical Disclosure Control - Version 1.1. A [Eurostat] Centre of Excellence for Statistical Disclosure Control. [http://neon.vb.cbs.nl/cenex/CENEX-SDC\\_Handbook.pdf](http://neon.vb.cbs.nl/cenex/CENEX-SDC_Handbook.pdf) (2009). Accessed 15 October 2009.
- [13] Koch, I., Meinken, H.: The Employment Panel of the German Federal Employment Agency" *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*. 124(2), 315 – 325 (2004)
- [14] Kohlmann, A.: The Research Data Centre of the Federal Employment Service in the Institute for Employment Research. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*. 125(3), 437 – 445 (2005)
- [15] Kölling, A.: The IAB-Establishment Panel. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*. 120(2), 291 – 300 (2000)
- [16] Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI): Wege zu einer besseren informationellen Infrastruktur" Nomos, Baden-Baden (2000)
- [17] Jacobebbinghaus, P., Seth, S.: The German integrated employment biographies sample IEBS. *Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften*. 127(2), 335 – 342 (2007)
- [18] Meßmann, S., Bender, S., Rudolph, H., Hirseland, A., Bruckmeier, K., Wübbecke, C., Dundler, A., Städele, D., Schels, B.: Lebenssituation und Soziale Sicherung 2005 (LSS 2005) - IAB-Querschnittsbefragung SGB II. Handbuch-Version 1.0.0. FDZ Datenreport. 04/2008 (2008)
- [19] Oertel, M., Schneider, A., Zimmermann, R.: Kundenbefragung zur Analyse der Organisationsstrukturen in der Grundsicherung nach SGB II - Dokumentation der Scientific-Use-Files. FDZ Datenreport. 04/2009 (2009)

- [20] Schneider, H.: Mehr und bessere Daten für die Arbeitsmarktforschung. RatSWD Working Paper Series. Working Paper No. 106 (2009)
- [21] Schneider, H., Wolf, C.: Die Datenservicezentren als Teil der informationellen Infrastruktur. In: Rolf, G., Zwick, M., Wagner, G. (eds.) Fortschritte der informationellen Datenstruktur in Deutschland. pp. 237 – 249. Nomos, Baden-Baden, (2008)
- [22] Spengler, A.: The Establishment History Panel. Schmollers Jahrbuch. Zeitschrift für Wirtschafts- und Sozialwissenschaften. 128(3), 501 – 509 (2008)
- [23] Trappmann, M., Achatz, J., Christoph, B., Wenzig, C.: PASS: A new panel study for labour market research. International Journal of Manpower. 30(7) (2009a) *forthcoming*.
- [24] Trappmann, M., Christoph, B., Achatz, J., Wenzig, C., Müller, G., Gebhardt, D.: Design and stratification of PASS - a new panel study for research on long term unemployment. IAB Discussion Paper. 05/2009 (2009b)
- [25] Wissenschaftsrat: Stellungnahme zum Institut für Arbeitsmarkt- und Berufsforschung (IAB). Nürnberg. Drs. 8175-07 (2007)